

# 母音識別の方法 (I)

藤 宗 寛 治<sup>\*</sup>・村 本 浩<sup>\*\*</sup>・岩 原 正 吉<sup>\*\*\*</sup>

## A Method on The Identification of Japanese Vowels. (I)

Hiroharu FUJISO Hiroshi MURAMOTO Masayoshi IWAHARA

(Received 18 Oct, 1967)

Recently, the recognition of patterns in figures and vocal sounds has been powerfully developed by large, high speed computers.

But, on the identification of vocal sounds, we have not heard of any satisfactory results yet.

This paper briefly describes our test results on the identification of Japanese vowels with regard to frequency spectrum by the computer, FACOM-270-20.

### 1 緒 言

生体系の情報処理に関する研究のうち、図形認識および音声認識については、大容量・高速度型電子計算機の使用が容易になるに従って近年多くの成果が上げられて来た。とくに図形認識のうち文字識別などに非常な成果が見られる。

それに対して、音声識別は従来、主として、アナログ的な手法によっているために文字識別の場合のような良い結果が得られていない。

筆者らは音声識別における判定機構として電子計算機を利用する方法を考え、母音の判定について実験を試みたところ、所期の結果を得たので、これを報告する。

### 2 母音の特徴

われわれ、人間の音声は音韻学的に、母音、子音、無声子音の三つに大別され、そのうち、子音および無声子音はさらに摩擦音、破裂音、鼻音、半田音、破

擦音等に分類されることが知られている。これら音声は、先行子音や破裂性の子音を除けば、3～4 Kc/sの帯域の内にある。しかし、それらの音声スペクトルはこの帯域内に一様に分布しているのではなくて、ある限られた周波数領域に特徴的な集合として存在している。このことはVOCODERを用いたソナグラムや時間的に平均化した音声の周波数分析によって明らかであるところである。

これらの結果によって、母音は極めて明確な三つのスペクトル集合によって特徴づけられていることが知られている。

その特徴的スペクトル集合の周波数軸上の相対的位置は各個人によって多少の変移がある。それらはそれぞれ低周波数領域から第1フォルマント、第2フォルマント、第3フォルマントと名付けられている。それらフォルマントの平均的位置は表一<sup>1)</sup>のとおりである。

表一の結果は音声合成について中田<sup>2)</sup>の行なった実験ともよく一致している。

表一 日本語母音のフォルマント

(男声)	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	(女声)	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>
ア	600—800c/s	1000—1400c/s	2700—3100c/s	ア	1000—1200c/s	1600—2000c/s	2300—3200c/s
エ	350—550	1500—2000	2500—3000	エ	400—600		2200—3000
イ	250—350		2400—3000	イ	350—450		2800—3300
オ	420—500	760—1000	1300—2000	オ	500—650	800—1000	1600—3000
ウ	300—480	1000—1400	2000—3000	ウ	400—500	1500—1800	2500—3000

F<sub>1</sub>, F<sub>2</sub>, F<sub>3</sub> はそれぞれフォルマントのある所の範囲を示す。

\* 教授 \*\* 助教授 \*\*\* 助手

上述のように母音のフォルマントは比較的安定したものであるから、この特性を利用すれば、母音の識別は容易かつ可能となる。

母音の識別についての研究<sup>87, 92</sup>の多くは、これらフォルマントに注目して行なわれ、中田<sup>87, 92, 100</sup>らは短時間周波数スペクトル分析の方法、坂井<sup>119</sup>らは零交叉波分析の方法などによっているが、いずれもアナログ的な取り扱いが主体であって、全面的にデジタル計算機によったものではない。

筆者らの研究はこのフォルマントについて母音の決定・判別を全面的にデジタル電子計算機によって行なったものである。

### 3 線形識別関数<sup>120, 121</sup>の導入

上述の母音の特徴をさらに処理し易くするために線形識別関数を導入する。この場合、以下の諸事項はすでに満足されているものと仮定する。

(1) 日本語の単音節母音であること。

(2) 母音音節はVOCODERなどによって、すでに周波数スペクトルに分析され、符号化されていること。

(3) ピッチ抽出により母音区間の判定はアナログ的に処理されていること。

(4) 分析されたスペクトルは時間的に平均化され、一時的に磁気テープなどに記憶されていること。

(5) 必要な情報量は符号化されたものの中にすべて含まれていること。

以上のことから、この実験は電子計算機のデータ入力点から開始されることになる。先ず、入力音声の波形はアナログ量の段階で振幅について正規化を行ないさらにデジタルの段階でもう一度厳密に正規化する。

以上の前提操作はすでに行ってあるものとして識別関数を考慮するものとする。

前述したように、母音はそのフォルマントによって特徴づけられるから、これらフォルマントを分離・抽出する方法を考えれば良い。ここでは、中田らによって確認されたフォルマントの標準的なパターンを電子計算機に記憶させ、母音入力と比較させることにした。

しかし、フォルマントは各個人によって標準的なパターンに対して相対的な位置のずれおよびパターンのひずみを示すから、この場合は入力はそのままの形では判定せずに、先ず、標準パターンからの位置のずれを調べ、そのずれの程度に応じた重みづけを行なってさらに変形されたパターンとし、それによって母音を

識別するが、この場合、同時にフォルマントの分離機能をも持たせ、識別率をあげるためにクラスター化された各フォルマントについてエネルギースペクトルの積分を行ない、それぞれの積分値がある設定された閾値を越えたか否かを判定して最終的に決定を行なう。

考慮した線形識別関数として、この場合、重みベクトルの集合をも与えるが、周波数に関連してある広がりを持つ正規分布関数を用いるものとする。

さらに、この線形識別関数に対して、次のような仮定をした。

i) 第1フォルマントのスペクトルには局在性が強く、エネルギー値も大きい。従って、判別上もっとも重要であり、重みベクトルの範囲を広く取ると誤判別の恐れが増すから、ベクトルの範囲をせまく、大いさは大きくとる。

ii) 第2フォルマントに対する重みベクトルの範囲は第1フォルマントに対するよりも広く、その大いさは小さく取る。

iii) 第3フォルマントに対する重みベクトルは、母音の特徴に対する寄与の程度から考えて、単にその範囲の存在の有無を調べるだけでも有効と考えられるが今回は与えるものとする。従って、将来、判別上に格別の寄与が認められぬ場合にはこれを取り除くものとする。

筆者らの用いた線形識別関数を下に示そう。

$$y = 1/\sqrt{2\pi\sigma} \cdot E \times P\{(x-m)/2\sigma^2\}$$

ここで、 $y$  は重みベクトルの各々の値を与える。 $\sigma$  は重みベクトルの広がり範囲を与え、 $m$  は重みベクトルの中心位置を与える。また、 $x$  は入力音声の分析された周波数となる。

更に、クラスター化は次によって行なう。

$$(\omega \cdot x) = \sum_{i=1}^n x_i \cdot \omega_i$$

この時、

$$\omega(i), x(i) \text{ ならば, } x'_i = x_i \cdot \omega_i$$

$$\omega(i), x(j) \text{ ならば, } x'_j = x_j \cdot 0$$

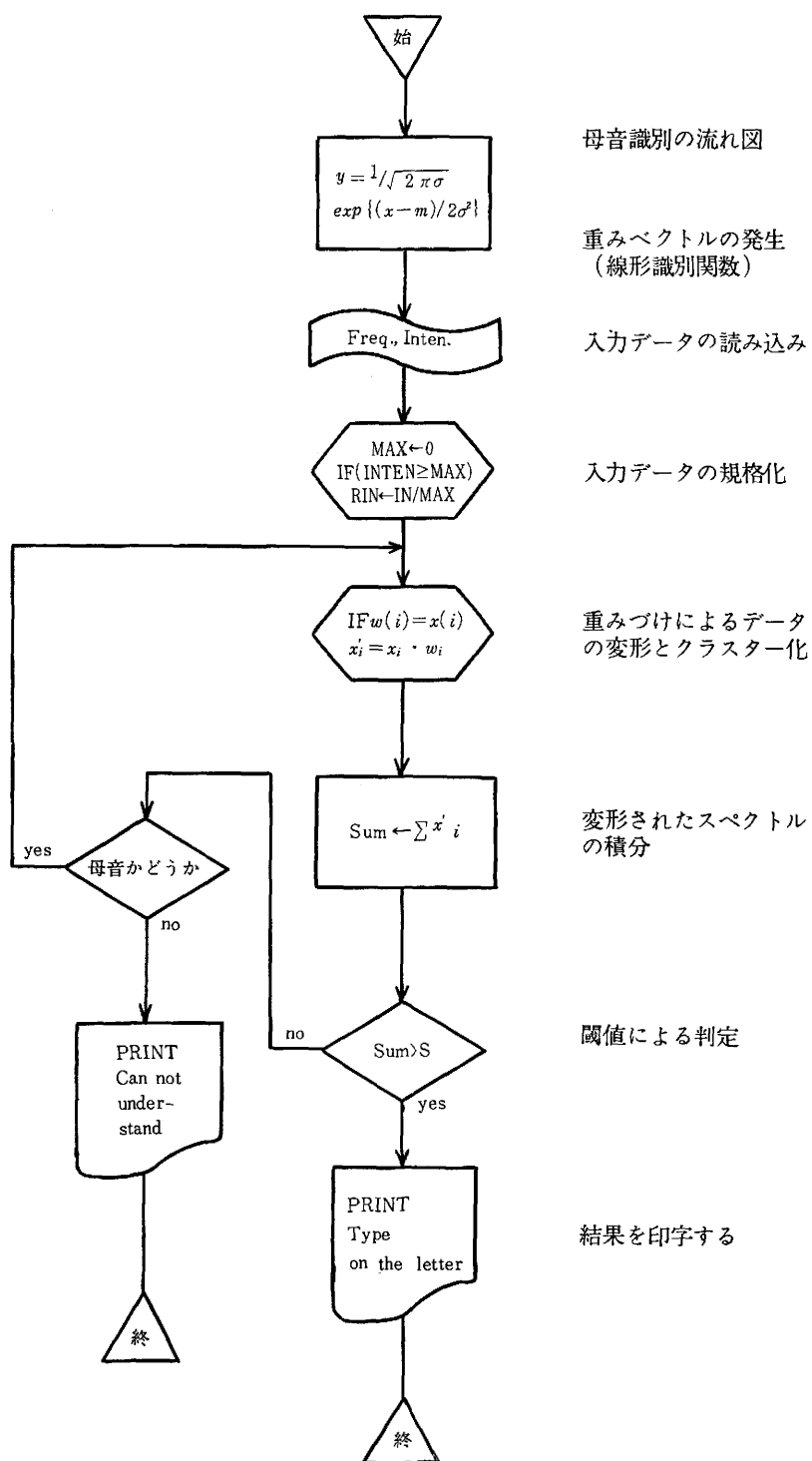
となるものとする。

ここで、 $\omega$  は重みベクトルであり、 $\omega_i$  はその大いさである。また  $x$  は観測値、ここでは音声入力値の集合であり、 $x_i$  はその大いさであり、 $x'_i$  は変形された入力値の成分である。

上式によって変形されたスペクトルの集合のそれぞれについて積分を行ない、閾値  $S$  について判定を与える。すなわち、

$$\text{Sum}(F_{1,2,3}) = \sum_{K1, K2, K3}^{n1, n2, n3} x'_{K1, K2, K3}$$

図 1



$$\text{Sum}(F_{1,2,3}) \geq S_{1,2,3} \rightarrow \text{YES}$$

となる。

ここで、Sumは積分の値であり、そのかっこ内のパラメータ $F_{1,2,3}$ は $F_1$ ,  $F_2$ ,  $F_3$ （各フォルマントを省略して書いたもの）の個々について上述の操作を行なうことを意味し、かつ $F_1$ ,  $F_2$ ,  $F_3$ についてAND操作をすることを意味する。

#### 4 プログラムについて

3で要求されたことに対して、FACOM—270—20に用いられるプログラムを作成した。

プログラムの流れ図の概要は第1図のとおりである。識別の際には母音5種類のすべてについて、そのうちのどれであるかを判定し、不明なものに対してはType Writer 上に Cannot understand と印字し、判定して識別されたものについては、判定された母音の名前によって、Type —A— on the letter のように印字する。

プログラム上特に問題になる点はないが、この実験では重みベクトルを電子計算機によって発生させたため所要時間が大となったので、実際には値を求めて置いて、予め電子計算機の記憶装置に格納して置くほうが有利である。また、この実験では全行程のプログラムを作成し、電子計算機を運転する予定であったが、操作の都合で分割して処理した。

#### 5 実験結果ならびに考察

電子計算機によって母音を判別させたプログラムの結果の一例を図2、図3、図4および図5に示す。図はそれぞれ、識別と分離のための重みベクトルの表、入力値、変形された人力の値およびそれら入力に対する答えを示す。

ところで、この実験では判定に相当の時間を要した。

それらは入力データや変形されたデータや重みベクトルの表などの入出力機構にかかる所要時間が殆んどで、勿論基本的にはこれらの時間は大幅に縮小が可能であるが、判定演算のみでも実時間動作として実用に供するのにはなお将来の検討を要する。

そのためには、判定のためのパラメータの選択について、情報量をへらして、かつその特徴を失わないための方策を見出すことが必要である。

このことについては、つぎの機会にさらに検討を加えてみたい。

#### 6 結 言

この研究は母音の識別に電子計算機を全面的に活用する方法について検討するため、プログラムその他について研究を行なったもので、その一部については所期の結果を得ることができたが、なお詳細にわたって

図2 重みベクトルの一例

WEIGHTING FACTOR TABLE ON VOWEL SOUND, A.

KC	RATIO	KC	RATIO	KC	RATIO	KC	RATIO	KC	RATIO
0.709	0.0230	0.719	0.0267	0.729	0.0310	0.739	0.0358	0.750	0.0412
0.759	0.0472	0.769	0.0541	0.779	0.0617	0.789	0.0701	0.799	0.0795
0.809	0.0898	0.819	0.1011	0.829	0.1136	0.839	0.1271	0.849	0.1418
0.859	0.1578	0.869	0.1750	0.879	0.1935	0.889	0.2133	0.899	0.2343
0.909	0.2567	0.919	0.2803	0.929	0.3051	0.939	0.3311	0.949	0.3583
0.959	0.3864	0.969	0.4155	0.979	0.4454	0.989	0.4760	1.000	0.5071
1.009	0.5385	1.019	0.5702	1.029	0.6018	1.039	0.6333	1.049	0.6643
1.059	0.6947	1.069	0.7243	1.079	0.7528	1.089	0.7800	1.099	0.8057
1.109	0.8297	1.119	0.8517	1.129	0.8717	1.139	0.8894	1.149	0.9046
1.159	0.9172	1.169	0.9272	1.179	0.9344	1.189	0.9387	1.199	0.9402
1.209	0.9387	1.219	0.9344	1.229	0.9272	1.239	0.9172	1.250	0.9046
1.259	0.8894	1.269	0.8717	1.279	0.8517	1.289	0.8297	1.299	0.8057
1.309	0.7800	1.319	0.7528	1.329	0.7243	1.339	0.6947	1.349	0.6643
1.359	0.6333	1.369	0.6018	1.379	0.5702	1.389	0.5385	1.399	0.5071
1.409	0.4760	1.419	0.4454	1.429	0.4155	1.439	0.3864	1.449	0.3583
1.459	0.3311	1.469	0.3051	1.479	0.2803	1.489	0.2567	1.500	0.2343
1.509	0.2133	1.519	0.1935	1.529	0.1750	1.539	0.1578	1.549	0.1418
1.559	0.1271	1.569	0.1136	1.579	0.1011	1.589	0.0898	1.599	0.0795
1.609	0.0701	1.619	0.0617	1.629	0.0541	1.639	0.0472	1.649	0.0412
1.659	0.0358	1.669	0.0310	1.679	0.0267	1.689	0.0230	1.699	0.0197

図3 入力スペクトルの一例 INPUT FREQUENCY(KC)/INPUT LEVEL(DB) (A)

KC	DB	KC	DB	KC	DB	KC	DB	KC	DB
0.000	0.0000	0.500	0.1859	0.509	0.3239	0.519	0.5399	0.529	0.8819
0.539	1.3949	0.549	2.1449	0.559	3.2099	0.569	4.6709	0.579	6.6119
0.589	9.0989	0.599	12.1829	0.609	15.8609	0.619	20.0849	0.629	24.7379
0.639	29.6339	0.649	34.5239	0.659	39.1229	0.669	43.1159	0.679	46.2179
0.689	48.1859	0.699	48.8579	0.709	48.1859	0.719	46.2179	0.729	43.1189
0.739	39.1229	0.750	34.5269	0.759	29.6339	0.769	24.7379	0.779	20.0849
0.789	15.8609	0.799	12.1829	0.809	9.0989	0.819	6.6119	0.829	4.6709
0.839	3.2099	0.849	2.1449	0.859	1.3949	0.869	0.8819	0.879	0.5399
0.889	0.3239	0.899	0.1859						

INPUT FREQUENCY(KC)/INPUT LEVEL(DB) (B)									
KC	DB	KC	DB	KC	DB	KC	DB	KC	DB
3.140	51.0119	0.719	0.8039	0.739	1.0769	0.759	1.4189	0.779	1.8539
0.799	2.3879	0.819	3.0359	0.839	3.8159	0.859	4.7369	0.879	5.8079
0.899	7.0319	0.919	8.4119	0.939	9.9359	0.959	11.5949	0.979	13.3649
1.000	15.2159	1.019	17.1089	1.039	19.0019	1.059	20.8439	1.079	22.5869
1.099	24.1739	1.119	25.5539	1.139	26.6849	1.159	27.5189	1.179	28.0349
1.199	28.2089	1.219	28.0349	1.239	27.5189	1.259	26.6849	1.279	25.5539
1.299	24.1739	1.319	22.5869	1.339	20.8439	1.359	19.0019	1.379	17.1089
1.399	15.2159	1.419	13.3649	1.439	11.5949	1.459	9.9359	1.479	8.4119
1.500	7.0319	1.519	5.8079	1.539	4.7369	1.559	3.8159	1.579	3.0359
1.599	2.3879	1.619	1.8539	1.639	1.4189	1.659	1.0769	1.679	0.8039
1.699	0.5939								

図4 変形された入力スペクトル

INPUT FREQUENCY(KC)/REDUCED INPUT LEVEL(DB) (A)									
KC	DB	KC	DB	KC	DB	KC	DB	KC	DB
0.000	0.0000	0.500	0.0011	0.509	0.0034	0.519	0.0097	0.529	0.0259
0.539	0.0648	0.549	0.1533	0.559	0.3434	0.569	0.7272	0.579	1.4572
0.589	2.7597	0.599	4.9475	0.609	8.3857	0.619	13.4469	0.629	20.3989
0.639	29.2724	0.649	39.7301	0.659	51.0202	0.669	61.9662	0.679	71.2034
0.689	77.3962	0.699	79.5700	0.709	77.3962	0.719	71.2034	0.729	61.9748
0.739	51.0202	0.750	39.7371	0.759	29.2724	0.769	20.3989	0.779	13.4469
0.789	8.3857	0.799	4.9475	0.809	2.7597	0.819	1.4572	0.829	0.7272
0.839	0.3434	0.849	0.1533	0.859	0.0648	0.869	0.0259	0.879	0.0097
0.889	0.0034	0.899	0.0011						

INPUT FREQUENCY(KC)/REDUCED INPUT LEVEL(DB) (B)									
KC	DB	KC	DB	KC	DB	KC	DB	KC	DB
3.140	0.0000	0.719	0.0215	0.739	0.0386	0.759	0.0671	0.779	0.1145
0.799	0.1900	0.819	0.3072	0.839	0.4853	0.859	0.7479	0.879	1.1244
0.899	1.6483	0.919	2.3587	0.939	3.2908	0.959	4.4814	0.979	5.9541
1.000	7.7175	1.019	9.7572	1.039	12.0358	1.059	14.4824	1.079	17.0057
1.099	19.4794	1.119	21.7668	1.139	23.7362	1.159	25.2431	1.179	26.1986
1.199	26.5249	1.219	26.1986	1.239	25.2431	1.259	23.7362	1.279	21.7668
1.299	19.4794	1.319	17.0057	1.339	14.4824	1.359	12.0358	1.379	9.7572
1.399	7.7175	1.419	5.9541	1.439	4.4814	1.459	3.2908	1.479	2.3587
1.500	1.6483	1.519	1.1244	1.539	0.7479	1.559	0.4853	1.579	0.3072
1.599	0.1900	1.619	0.1145	1.639	0.0671	1.659	0.0386	1.679	0.0215
1.699	0.0117								

図5 判定条件に対する Type writer 上への結果印字

TOTAL 1 =	846.21118	TOTAL 1 =	423.10278
TOTAL 2 =	755.49389	TOTAL 2 =	423.04125
TOTAL 3 =	613.66772	TOTAL 3 =	420.71997
TYPE -A- ON THE LETTER		CAN NOT UNDERSTAND	
(A)		(B)	

は実際の母音入力波に対してこの方法を用いて、その結果から更に帰納的に識別率を高くするための識別関数の改良やパラメータ処理などについて改善工夫の要があることが認められる。

#### 文 献

- 1) 小幡重一；“実験音響学”，岩波書店，(36年10月)
- 2) 中田和男；“日本語音声の合成的研究”，電波研究所季報，(61年12月)
- 3) 鈴木誠史，中田和男；“母音の識別”，電波研究所季報，7，163，(1961)
- 4) 鈴木誠史，中津井護；“ホルマント周波数パターンによる単音節の分析と識別”電波研究所季報，9，281，(1963)
- 5) 鈴木久喜，高橋敬，大泉充郎；“母音認識の方法について”電気音響学会専門委員会資料，(59年9月)
- 6) 鈴木久喜，大泉充郎；“線形判別関数を用いた母音識別”，電気通信学会誌，47，315，(1964)
- 7) 坂井利之，堂下修司；“会話音声識別の一方式”，オートマトンと自動制御研究専門委員会資料，(62年1月)
- 8) 鈴木誠史，中田和男，前園研一；“日本語単音節の分類と識別”，電波研究所季報，9，124，(1963)
- 9) 鈴木誠史，角川靖夫，中田和男；“モーメント計算によるホルマント周波数の抽出”，電波研究所季報，9，7，(1963)
- 10) 中田和男；“音声の識別”，電気通信学会誌，46，291，(1963)
- 11) 坂井利之，堂下修司，橋本清；“日本語単音節の符号化”，オートマトンと自動制御専門委員会資料，(61年1月)
- 12) 坂井利之 他；“パターン認識の理論”，共立出版株式会社，(67年5月)
- 13) 飯島泰蔵；“パターン認識”，電気通信学会誌，46，1582，(1963)
- 14) 宇田川銑久，福村晃夫，阿部圭一；“認識パラメータ集合の動的計画法を用いた一構成法”，電気通信学会誌，48，926，(1965)
- 15) A. G. Arkadev & E. M. Braverman；“Teaching Computers to Recognize Patterns”，ACADEMIC PRESS，(67年8月)

(昭和42年10月18日受理)